symmetry

MDPI

Article

A Digital Twin-Based Speaker Placement Planning Tool for Indoor Environments

Zhikang Li, Nobuo Funabiki *, Kadek Suarjuna Batubulan, I Nyoman Darma Kotama, Putu Sugiartawan and Anak Agung Surya Pradhana

Department of Information and Communication Systems, Okayama University, Okayama 700-8530, Japan; p8d81de9@s.okayama-u.ac.jp (Z.L.); pzc37um1@s.okayama-u.ac.jp (K.S.B.); p9363bg2@s.okayama-u.ac.jp (I.N.D.K.); p18z9yov@s.okayama-u.ac.jp (P.S.); p44c722y@s.okayama-u.ac.jp (A.A.S.P.)

* Correspondence: funabi-n@cc.okayama-u.ac.jp

Abstract

Nowadays, speakers are essential components for message delivery in indoor environments, including lectures, public addresses, and emergency announcements. Their physical placement should ensure adequate direct-sound audibility across occupant service areas while maintaining installation feasibility. A digital twin is a technology that allows an infrastructure layout to be designed and evaluated virtually on a computer before physical installation by reconstructing an indoor environment as a 3D model. In previous studies, we have proposed a method to reconstruct a 3D indoor model of an indoor environment from its 360° panoramic images using 3D Gaussian Splatting (3DGS) and a 3D point cloud, and applied it to surveillance camera placement. In this paper, we propose a digital twin-based speaker placement planning tool for indoor environments by generalizing the previous method to direct-sound acoustic simulation. This tool consists of four stages: (1) reconstructing a 3D indoor model from 360° panoramic images and extracting floor and desk receiver surfaces, (2) assigning the initial speaker budget based on the reconstructed floor area, (3) determining speaker mounting positions on valid ceiling regions using *K*-means spatial clustering under obstacle and boundary constraints, and (4) simulating broadband direct-sound sound pressure level (SPL) across floor and desk receiver surfaces. For evaluation, the proposed tool was deployed across three real-world indoor scenarios: a basketball hall (32.15 m × 21.99 m), a furnished office (7.17 m × 6.17 m), and a non-convex L-shaped office (23.00 m^2). The experimental results showed that in each scenario, the generated layout achieved complete direct-sound audibility compliance across all sampled physical test locations (≥60 dB floor/≥65 dB desk) with zero detected hotspots exceeding 85 dB, maintaining SPL values between 66.91 dB and 77.88 dB. An on-site physical measurement campaign confirmed that the generated layouts satisfy target audibility thresholds under real room conditions, with mean absolute errors between 1.78 dB and 2.41 dB. These results confirm the practical utility of our approach as an initial geometry-driven planning tool for indoor audio infrastructure deployment.

Keywords: digital twin; panoramic image; 3D Gaussian splatting; 3D point cloud; speaker placement planning; indoor environment; sound pressure level (SPL); acoustic digital twin

Check for updates

Academic Editor: Jie Yang

Received: 6 August 2026
Revised: 10 September 2026
Accepted: 15 September 2026
Published: 17 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Nowadays, speakers are widely used for message delivery in indoor environments, including for lectures, talks, background audio, and emergency announcements. In any

use, sufficient sound audibility and appropriate pressure levels must be ensured across the intended occupant service areas. In an indoor environment, sound distribution is affected by propagation distances from speakers, the room geometry, surface conditions of the floor, walls, and furniture, and positions and directions of speakers. Unsuitable placements of speakers can produce insufficient sound levels in remote corners or excessive, uncomfortable sound at specific locations. Speaker placement is also often determined through manual inspection or repeated physical testing, which requires time and makes placement comparison difficult. Speaker placement can be planned and evaluated by simulating a sound propagation model in an indoor environment before physical installation. In ref. [1], Morales et al. determined a receiver location that maximizes the Speech Transmission Index (STI) using an indoor 3D model. In ref. [2], Delabie et al. presented a simulation framework that supports multiple speakers, directivity, image-source calculation, ray tracing, air absorption, and reverberation. These studies show that virtual acoustic analysis using a 3D model can support indoor sound planning. On 3D models for acoustic analysis, several works have been reported. In ref. [3], Marković et al. converted 3D point-cloud data into 3D room models for acoustic simulation. In ref. [4], Bergerot et al. used a point cloud to display architectural information and acoustic descriptors. Digital twin (DT) technology can further connect physical environments with digital models for simulation, monitoring, assessment, and planning [5,6]. In our previous study, we presented a 3D indoor modeling method using 360° panoramic images and its application to CCTV camera placement [7]. This digital twin-based camera placement approach extracts the floor, desks, ceiling, room boundary, and obstacle regions, and places the cameras across valid walls using *K*-means clustering on the 3D model generated from panoramic images. The visual coverage of the floor was virtually evaluated through ray casting on the 3D model. However, existing methodologies treat panoramic 3D reconstruction, loudspeaker spatial positioning, and room acoustic simulation as disconnected tasks, lacking an integrated, automated image-to-placement pipeline within a unified digital twin framework. In this paper, we propose a digital twin-based speaker placement planning tool for indoor environments by generalizing the previous method to direct-sound acoustic simulation. This tool consists of four stages: (1) capturing 360° panoramic images of the target environment and reconstructing the 3D indoor model using 3D Gaussian Splatting (3DGS) and point cloud filtering, (2) estimating the floor area and assigning an initial speaker budget based on an area-based allocation rule, (3) determining candidate ceiling speaker positions using *K*-means spatial clustering constrained by reconstructed room boundaries and obstacles, and (4) simulating broadband sound pressure level (SPL) to verify direct sound coverage across floor and desk receiver surfaces and visualizing them on the 3D model. For evaluation, we applied the proposed tool to three real-world scenarios with different room layouts: a basketball hall (32.15 m × 21.99 m), a furnished office (7.17 m × 6.17 m), and a non-convex L-shaped office (23.00 m^2). Experimental results showed that in each scenario, the obtained placement achieved complete direct-sound audibility compliance across all sampled physical test locations ($\geq$60 dB floor/$\geq$65 dB desk) with zero detected hotspots exceeding 85 dB, maintaining direct-sound SPL values between 66.91 dB and 77.88 dB. In addition, an on-site physical measurement campaign confirmed that the generated layouts achieve full audibility compliance at the sampled receiver points under real physical conditions. These results confirm the practical utility of our approach for digital twin-based audio infrastructure planning.

The rest of this paper is organized as follows: Section 2 reviews related works. Section 3 discusses adopted hardware and software technologies. Section 4 presents the proposed system architecture and methodology. Section 5 details the system implementation. Section 6 evaluates the performance, baseline comparisons, and physical validation across three

rooms. Section 7 discusses runtime performance, scaling boundaries, and physical acoustic scope. Finally, Section 8 concludes this paper with directions for future work.

2. Related Works

In this section, we review related works across digital twin building applications, indoor acoustic prediction, loudspeaker placement optimization, and panoramic indoor reconstruction.

2.1. Digital Twin-Based Building Applications

Digital twin technology has been applied to building operation, facility management, and occupancy analysis. These applications use virtual building models and real-time sensor data to represent the conditions of physical indoor environments [5,6].

In ref. [8], Seghezzi et al. presented an occupancy-oriented digital twin for facility management. They used camera-based sensors to collect occupancy data and connect the data with a virtual building model. However, their study focuses on occupancy monitoring and does not consider sound propagation or speaker placement.

In ref. [9], Cairoli et al. presented a digital twin for acoustic and stage facility management in a multipurpose hall. Their method connects building information, acoustic properties, and different hall configurations. However, the system is used to manage existing hall configurations and does not automatically search for speaker locations.

In ref. [10], Shahzad et al. investigated the characteristics and applications of digital twin technology for built assets. They discussed connections between building information models, sensor data, and physical building operation. However, their study focuses on general building applications and does not provide a procedure for indoor audio coverage planning.

In ref. [11], Aljamal et al. proposed a renewable energy-aware carbon digital twin for sustainable construction operations using AI-driven multi-objective optimization. However, their framework focuses on construction scheduling rather than indoor acoustic coverage planning.

In ref. [12], Lu et al. developed a digital twin framework integrated with TSMixer for time-series remaining useful life prediction of mechanical bearings. However, their study focuses on industrial prognostic health management and does not address 3D indoor spatial reconstruction or speaker placement.

2.2. Indoor Acoustic Prediction

Indoor acoustic prediction methods have been used to estimate sound behavior under different room shapes, surface properties, and obstacle conditions. The methods allow acoustic conditions to be examined before physical installation.

In ref. [13], Forssén et al. presented a hybrid room acoustic prediction tool for open-plan offices. Their method combines phased image sources and acoustic radiosity to calculate reflections and shielding effects of office screens. This method assumes simple room shapes and does not determine speaker locations.

In ref. [14], de Bort et al. presented a graphical method for examining multiple sound reflections across an audience area. Their method uses ray-tracing results to support the selection of absorbers and reflectors for improving speech clarity. However, their study focuses on room acoustic treatment and does not consider speaker placements from reconstructed indoor geometry.

In ref. [15], Zarei et al. examined the spatial uniformity of a sound-masking system in an open-plan office. They evaluated how the number of loudspeakers, partition height, and ceiling properties affect the sound field. Their study only examines prepared loudspeaker configurations and does not generate speaker placement.

In ref. [16], Scheibler et al. presented Pyroomacoustics , a Python (version 3.12) package for room acoustic simulation and array processing. Along with particle-tracing numerical simulation tools such as SPPS [17], this package supports room impulse response generation using the image-source method. However, these tools do not provide an integrated procedure for 3D indoor reconstruction and automated speaker placement selection.

2.3. Loudspeaker Placement Optimization

Several studies have investigated loudspeaker placement for producing the required sound field in selected listening zones. These methods commonly select loudspeaker locations from a predefined candidate set.

In ref. [18], Zhao et al. proposed an iterative loudspeaker selection method for multizone sound field reproduction. Their method selects loudspeaker locations based on the orthogonal relation between candidate and selected loudspeakers. The candidate locations and acoustic transfer functions must be prepared before the selection process.

In ref. [19], Wen et al. proposed a loudspeaker placement method based on modal-domain analysis. Their method uses constrained matching pursuit to select loudspeaker locations for bright and dark sound zones. The method focuses on multizone sound field reproduction and does not reconstruct the target indoor environment from panoramic images.

2.4. Panoramic Indoor Reconstruction

Panoramic images have been used to estimate room boundaries and reconstruct indoor 3D models. These methods can generate room layouts without manually measuring the walls and corners in a room.

In ref. [20], Sun et al. proposed a method for reconstructing planar indoor structures from a single panoramic image. Their method separates horizontal and vertical planes to recover the main room surfaces. The reconstructed model focuses on planar structures and does not include speaker placement or sound coverage calculation.

In ref. [21], Jiang et al. proposed LGT-Net for indoor room layout estimation from a panoramic image. Their method uses geometric relations to estimate wall boundaries, room height, and corner positions. It estimates only the main room layout and does not consider furniture and other obstacles that may affect human-accessible regions.

In ref. [22], Wang et al. proposed PSMNet to estimate room layouts from two 360° panoramic images. Their method combines information from two viewpoints and improves layout estimation for large and complex rooms. The method is limited to room layout estimation and is not connected with acoustic prediction or speaker placement.

2.5. Comparative Summary and Acoustic Metric Justification

Table 1 summarizes the main characteristics of related indoor spatial and acoustic planning studies in comparison with our proposed digital twin tool across five key technical dimensions.

Broadband SPL vs. Speech Transmission Index (STI):

While speech intelligibility is comprehensively described by the Speech Transmission Index (STI) [1], computing STI requires obtaining the complete Room Impulse Response (RIR) and Modulation Transfer Function (MTF), which depend heavily on frequency-dependent surface absorption coefficients (α) and reverberation parameters (RT_{60}). In early-stage infrastructure planning, establishing adequate direct sound pressure level (SPL) is the fundamental prerequisite: sound signals must deliver sufficient acoustic pressure above ambient background noise ($\geq$60 dB floor/$\geq$65 dB desk) without exceeding discomfort thresholds (>85 dB). Therefore, broadband direct SPL serves as a robust, computationally direct audibility foundation, while STI modeling will be incorporated in future iterations that integrate surface material impedance.

Table 1. Comparison of related indoor acoustic/spatial planning studies and the proposed method.

Study	Input Data	Geometry Source	Placement Method	Acoustic Model	Validation
Morales et al. [1]	3D CAD mesh	Synthetic 3D model	Gradient descent (receiver search)	Wave & ray tracing (STI metric)	Numerical simulation
Delabie et al. [2]	Manual room dims	Parametric CAD	Fixed predefined grid	ISM & ray tracing	Simulated impulse response
Marković et al. [3]	LiDAR point cloud	Point-cloud mesh	Manual source/mic assignment	Geometrical acoustics	Simulated RT_{60}
Bergerot et al. [4]	Laser scans (TLS)	Point cloud/BIM	Fixed static sensor locations	Offline acoustic descriptors	Point-cloud visualization
Pradhana et al. [7]	360° panoramas	3D Gaussian Splatting	K-means on valid walls (CCTV)	Line-of-sight ray casting	Physical camera coverage
Zhao et al. [18]	Predefined array	Fixed boundary mesh	Orthogonal matching pursuit	Multi-zone transfer functions	Numerical simulation
Wen et al. [19]	Predefined array	Fixed boundary mesh	Modal-domain matching pursuit	Sound field reproduction	Numerical simulation
Proposed Tool	360° Panoramas	3D Gaussian Splatting + RANSAC	Area allocation + K-means on valid ceilings	Broadband direct-sound SPL	Pyroomacoustics + Physical SLM

3. Adopted Technologies

In this section, we introduce the hardware and software technologies used in the proposed speaker placement planning tool.

3.1. Hardware

3.1.1. Panoramic Camera

Insta360 X5 is used to capture 360° panoramic images of the target indoor environment [23]. Its dual fisheye lenses record the floor, ceiling, walls, and indoor objects from each capture position.

The captured spherical images are converted into equirectangular panoramic images. These images provide the visual data required for camera-pose alignment, 3D indoor reconstruction, and functional surface extraction.

3.1.2. Personal Computer

The program runs on a personal computer equipped with an Intel Core i9processor, 64 GB of RAM, and an NVIDIA GeForce RTX 4090 GPU with 24 GB of memory. The GPU is used during the 3D reconstruction process, while point-cloud processing, speaker placement, and SPL calculations are performed using Python-based software.

3.2. Software

3.2.1. Cupixvista and Open-Source Alternatives

Cupixvista processes the captured 360° equirectangular panoramic image sequence via visual SLAM to output: (1) estimated 6-DoF camera poses and trajectory, and (2) a dense, metric-scaled, colored 3D point cloud in standard `.ply` format [24]. Importantly, the downstream acoustic planning pipeline is fully decoupled from the photogrammetric front-end; open-source Structure-from-Motion (SfM) pipelines such as COLMAP or OpenSfM, combined with open-source 3D Gaussian Splatting (3DGS) implementations, can serve as direct open-source drop-in replacements without requiring proprietary software.

3.2.2. 3D Gaussian Splatting Framework

The 3D Gaussian Splatting (3DGS) method represents a scene using anisotropic three-dimensional Gaussian elements [25]. It is used to produce a detailed visual representation of the captured indoor environment.

Unlike conventional point-cloud photogrammetry, 3DGS models the scene as continuous 3D ellipsoids parameterized by spatial center $\mu = (x, y, z)$, 3D covariance $\mathbf{\Sigma}$, opacity $\sigma \in [0, 1]$, and spherical harmonics color coefficients. To obtain a geometric representation suitable for surface extraction, the Gaussian primitives are converted into a discrete point cloud by: (1) extracting spatial mean positions $\boldsymbol{\mu}$ of each Gaussian, (2) filtering out low-opacity primitives ($\sigma < 0.25$) and overly diffuse ellipsoids ($\max(\mathbf{s}) > 0.15$ m) to eliminate floating haze artifacts, and (3) applying Statistical Outlier Removal (SOR, $k = 20$, $\beta = 1.0$) and uniform 0.05 m voxel grid downsampling in Open3D.

3.2.3. *K*-Means Spatial Clustering

The *K*-means clustering algorithm (implemented via SciPy/scikit-learn [26] with *k*-means++ initialization) is utilized to partition spatial ceiling coordinates into a predefined number of clusters N determined by the room's floor area. The cluster centroids serve as candidate speaker mounting coordinates, which are subsequently checked against geometric boundaries and obstacle exclusion zones (detailed in Section 4).

3.2.4. Scientific Computing Stack

NumPy, *SciPy*, and Open3D are used for numerical calculation, spatial processing, and point-cloud analysis, respectively [27,28]. NumPy provides vectorized array operations for Euclidean distance calculations, SPL simulation, and logarithmic energy superposition. SciPy provides spatial indexing (KD-Trees) and geometric optimization functions. Open3D is used to read, process, and segment the reconstructed point cloud into floor, desk, ceiling, room-boundary, and obstacle regions.

3.2.5. Pyroomacoustics Reference Software

Pyroomacoustics is used as an independent reference simulator for direct-sound acoustic verification [16]. The simulation uses identical speaker coordinates, receiver grids, source power levels, and omnidirectional radiation assumptions. Reflection order zero ($N_{\text{order}} = 0$) is configured to simulate pure line-of-sight propagation, enabling numerical comparison against the proposed direct-sound SPL calculation.

4. System Architecture

In this section, we present the architecture of the proposed digital twin-based speaker placement planning tool for indoor environments.

4.1. Overview

The proposed tool follows a four-stage workflow, as illustrated in Figure 1. The stages are: (1) 3D indoor reconstruction and functional surface extraction, (2) area-based speaker allocation, (3) speaker placement using *K*-means clustering, and (4) direct-sound SPL calculation and result visualization.

First, 360° panoramic images are captured in the target indoor environment and processed to reconstruct its 3D representation. The reconstructed 3D model is then separated into floor, desk, ceiling, room-boundary, and obstacle regions. The floor and desk regions are used as receiver surfaces, while the ceiling, boundary, and obstacle regions are used to determine valid speaker mounting positions.

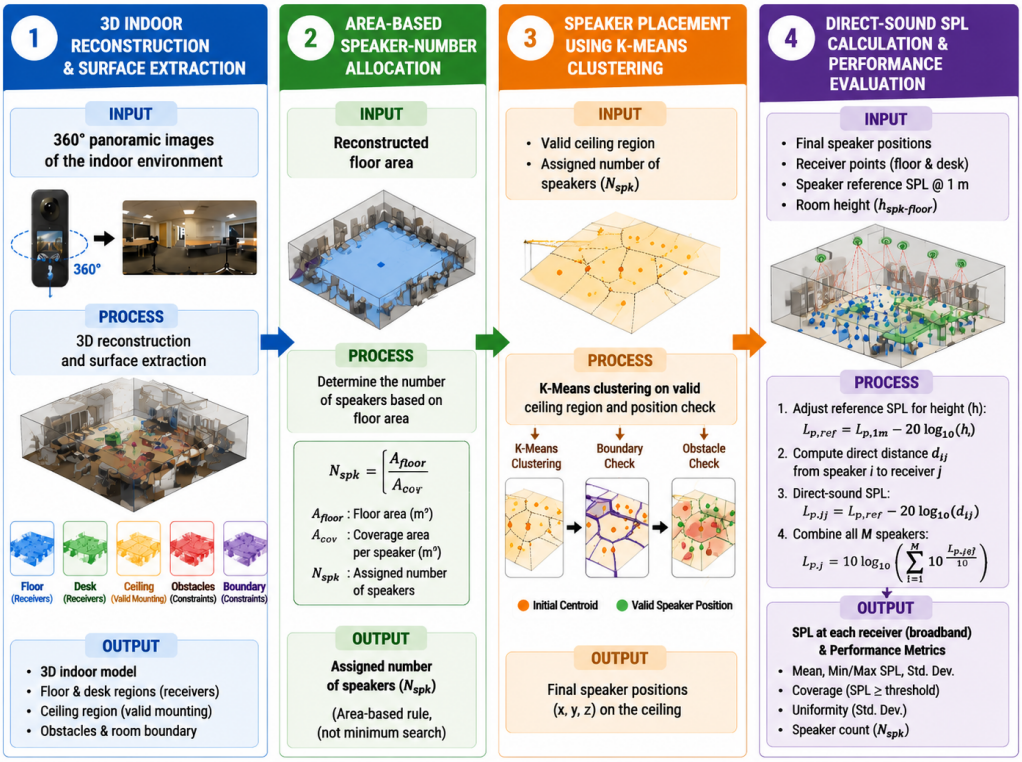

Figure 1. Workflow of the digital twin-based speaker placement planning tool.

Second, the initial number of speakers is assigned according to the reconstructed floor area. The current method applies a heuristic area-based allocation rule based on architectural acoustic coverage standards, serving as an initial budget rather than a mathematically unconstrained global minimum.

Third, the ceiling coordinates are divided into the assigned number of spatial clusters using the *K*-means clustering algorithm. The centroid of each cluster is selected as an initial speaker position. A constrained nearest-neighbor boundary and obstacle check is applied to project any invalid or colliding centroids onto valid ceiling coordinates.

Finally, on the 3D model, the broadband sound pressure level (SPL) value at each floor or desk receiver point is calculated. Each speaker is modeled as an acoustic source with fixed reference sensitivity and electrical power, evaluated via normalized spherical distance decay and energetic logarithmic superposition. The resulting direct SPL distribution is mapped onto the reconstructed 3D point cloud for visual inspection.

4.2. 3D Indoor Modeling and Surface Extraction Principles

The captured 360° panoramic images are processed using 3D Gaussian Splatting (3DGS) to produce a 3D representation of the target room. The reconstructed point data are then separated into floor, desk, ceiling, room-boundary, and obstacle regions.

The floor points, P_{floor}, represent general service areas. The desk points, P_{desk}, represent working surfaces that need a higher SPL value. The ceiling region is used for speaker placement, while the room-boundary and obstacle regions are used to identify infeasible speaker mounting positions. In the current model, obstacle regions are utilized strictly as spatial geometric constraints to prevent physical mounting collisions and do not attenuate direct sound rays.

4.3. Area-Based Speaker Allocation Formulation

The initial bounding floor area, A_{OBB}, is calculated from the horizontally oriented bounding box of the extracted floor region to establish the bounding envelope. In architec-

tural acoustic engineering guidelines (such as IEC 60849 [29]), the nominal direct-sound coverage radius of a ceiling loudspeaker with conical dispersion angle θ mounted at clearance $h_{\text{gap}} = h_{\text{ceil}} - h_{\text{ear}}$ is $r_{\text{cov}} = h_{\text{gap}} \tan(\theta/2)$. For typical indoor rooms ($h_{\text{gap}} \approx 1.8$–$2.2$ m) and standard $\theta \approx 90^\circ$ dispersion speakers, each unit effectively covers an overlapping area of $A_{\text{cell}} \approx 20$–$28$ m^2. Accordingly, the initial speaker budget N is assigned using the following stepped allocation rule:

$$N = \begin{cases} 2, & A_{\text{OBB}} < 45 \text{ m}^2, \\ 4, & 45 \leq A_{\text{OBB}} < 75 \text{ m}^2, \\ 6, & 75 \leq A_{\text{OBB}} < 120 \text{ m}^2, \\ 8, & A_{\text{OBB}} \geq 120 \text{ m}^2. \end{cases} \tag{1}$$

The selected number N is passed as the cluster count to the subsequent spatial placement stage. Note that while A_{OBB} is used to determine the initial cluster budget, the functional receiver areas reported in Section 6 represent exact concave mesh boundaries.

4.4. K-Means Spatial Speaker Placement and Shift Algorithm

Before spatial clustering, the valid ceiling candidate manifold $P_{\text{valid_ceil}}$ is formally extracted from the ceiling point cloud P_{ceiling} by enforcing boundary clearance and obstacle avoidance:

$$P_{\text{valid_ceil}} = \left\{ \mathbf{p} \in P_{\text{ceiling}} \;\middle|\; \text{dist}(\mathbf{p}, \partial\Omega_{\text{room}}) \geq \delta_{\text{wall}} \quad \text{and} \quad \min_{\mathbf{o} \in \mathcal{O}_{\text{obstacle}}} \|\mathbf{p}_{xy} - \mathbf{o}_{xy}\|_2 \geq \delta_{\text{obs}} \right\}, \tag{2}$$

where $\partial\Omega_{\text{room}}$ denotes the 2D bounding perimeter of the room with a wall clearance $\delta_{\text{wall}} = 0.30$ m, and $\mathcal{O}_{\text{obstacle}}$ denotes the 3D bounding boxes of ceiling fixtures (lighting, HVAC diffusers, beams) with a clearance buffer $\delta_{\text{obs}} = 0.40$ m.

The K-means algorithm partitions $P_{\text{valid_ceil}}$ into N clusters by minimizing within-cluster variance, yielding raw spatial centroids $\{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$. To guarantee full reproducibility, each centroid $\mathbf{c}_i$ is validated and adjusted via a constrained nearest-neighbor shift projection:

$$\mathbf{c}_i^* = \begin{cases} \mathbf{c}_i, & \text{if } \mathbf{c}_i \in P_{\text{valid_ceil}}, \\ \text{argmin}_{\mathbf{p} \in P_{\text{valid_ceil}}} \|\mathbf{p} - \mathbf{c}_i\|_2, & \text{otherwise}, \end{cases} \tag{3}$$

subject to the inter-speaker separation constraint $\|\mathbf{c}_i^* - \mathbf{c}_k^*\|_2 \geq d_{\min} = 1.5$ m ($\forall k \neq i$). The validated coordinates $\{\mathbf{c}_i^*\}_{i=1}^N$ define the active 3D speaker mounting positions.

4.5. Direct-Sound SPL Calculation

Each loudspeaker is modeled as an acoustic point source with nominal sensitivity S_0 (dB SPL @ 1 W/1 m) and electrical input power P_e (W). The reference sound pressure level at $d_0 = 1.0$ m is defined as:

$$L_{p,1\text{m}} = S_0 + 10 \log_{10}\left(\frac{P_e}{P_0}\right), \quad (P_0 = 1.0 \text{ W}). \tag{4}$$

In all simulations, standard commercial PA parameters ($S_0 = 90.0$ dB SPL, $P_e = 1.0$ W) are adopted, keeping the source power constant.

For speaker i at coordinate $\mathbf{c}_i^*$ and receiver point p_j, the direct-sound SPL contribution is calculated using dimensionally normalized spherical geometric spreading:

$$L_{p,i}(p_j) = L_{p,1\text{m}} - 20 \log_{10}\left(\frac{d_{ij}}{d_0}\right), \quad (d_0 = 1.0 \text{ m}), \tag{5}$$

where $d_{ij} = \|\mathbf{c}_i^* - p_j\|_2$ is the 3D Euclidean distance between speaker i and receiver point p_j.

Assuming uncorrelated broadband audio channels, the cumulative direct-sound SPL at receiver point p_j is obtained via energetic logarithmic superposition:

$$L_{p,\text{total}}(p_j) = 10 \log_{10} \left(\sum_{i=1}^{N} 10^{\frac{L_{p,i}(p_j)}{10}} \right). \tag{6}$$

The resulting value represents the broadband direct-sound SPL at each receiver point on the floor or a desk. The current model does not consider wall reflections, reverberation, material absorption, air absorption, speaker directivity, or obstacle-based sound blocking.

4.6. Result Visualization

The accepted speaker positions and calculated broadband SPL values are displayed on the reconstructed 3D indoor model. The visualization shows the relation among the room geometry, receiver regions, speaker positions, and direct-sound distribution. It supports visual inspection of the generated placement before a more detailed acoustic simulation or physical measurement is conducted.

5. Implementation

In this section, we present the implementation of the digital twin-based speaker placement planning tool. It follows the four-stage workflow described in Section 4.

5.1. Indoor Data Acquisition

First, the target indoor environment is captured using an Insta360 X5 panoramic camera mounted on an extended monopod stick. The operator selects several capture positions as they traverse the room to comprehensively record the entire 360° spatial environment, including the walls, floor, ceiling, and indoor furniture. Mounting the camera on an extended stick elevates the dual fisheye lenses, ensuring full spherical visual coverage of the room with minimal self-occlusion from the operator. Any incidental human presence during capture (such as the operator shown in Figure 2) is anonymized via facial blurring, with written informed consent obtained for publication.

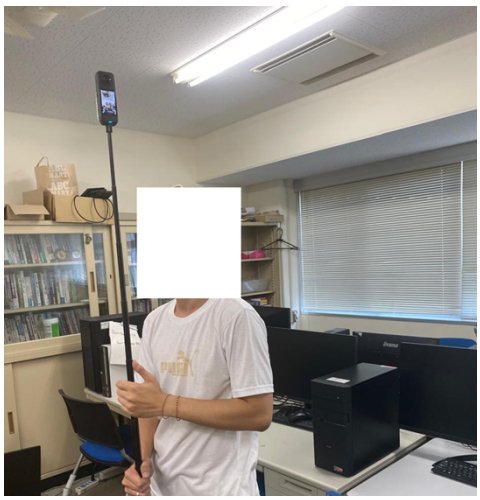

Figure 2. Indoor data acquisition using a 360° panoramic camera and the resulting equirectangular image.

We then convert the captured spherical images into equirectangular panoramic images for 3D reconstruction. Figure 2 shows the indoor data acquisition process and an example panoramic image.

5.2. 3D Reconstruction and Surface Segmentation Implementation

First, the panoramic images are aligned via Cupixvista and processed with 3D Gaussian Splatting to produce a dense 3D representation of the indoor environment. Figure 3 shows the reconstructed 3D model and the resulting extracted functional surfaces.

The filtered point cloud is segmented into functional architectural surfaces using sequential RANSAC plane fitting and normal filtering in Open3D, with the parameter configurations listed in Table 2:

1. Floor and Ceiling Plane Segmentation: RANSAC identifies dominant horizontal planes with surface normal vectors satisfying $|\mathbf{n} \cdot \hat{\mathbf{z}}| \geq 0.92$. The lowest detected horizontal plane is classified as floor (P_{floor}), and the highest is classified as ceiling (P_{ceiling}).
2. Desk Working Surfaces: Intermediate horizontal planes within the elevated workplane range ($z \in [z_{\min} + 0.70\,\text{m}, z_{\min} + 1.05\,\text{m}]$) with a surface area $> 0.5\,\text{m}^2$ are segmented as desk receiver surfaces (P_{desk}).
3. Room Perimeter and Obstacles: Vertical planar segments ($|\mathbf{n} \cdot \hat{\mathbf{z}}| \leq 0.15$) form the perimeter wall envelope $\partial\Omega_{\text{room}}$, while residual non-planar clustered points inside the room volume are classified as obstacle collision volumes $\mathcal{O}_{\text{obstacle}}$.

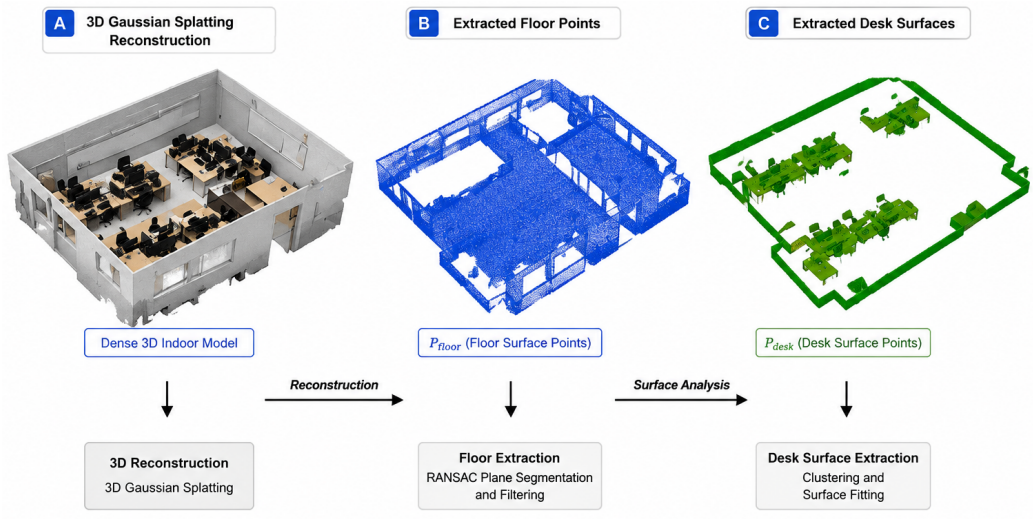

Figure 3. 3D indoor reconstruction and functional surface extraction: (**A**) reconstructed 3D model, (**B**) extracted floor points, and (**C**) extracted desk points.

Table 2. Parameters used for RANSAC functional surface segmentation.

Parameter	Setting
RANSAC distance threshold	0.03 m
RANSAC iterations	1000
Horizontal normal constraint ($\lvert\mathbf{n} \cdot \hat{\mathbf{z}}\rvert$)	$\geq$0.92 (angle $< 23°$)
Vertical normal constraint ($\lvert\mathbf{n} \cdot \hat{\mathbf{z}}\rvert$)	$\leq$0.15 (angle $< 8.6°$)
Desk height extraction window	$z_{\min} + 0.70$ m to $z_{\min} + 1.05$ m
Floor/Ceiling plane tolerance	$\pm$0.05 m

To evaluate segmentation fidelity, classification accuracy was tested against 2500 manually annotated ground-truth reference points across the experimental rooms. As summarized in Table 3, the segmentation achieved an overall accuracy of 96.4%.

Table 3. Functional surface segmentation accuracy against manual ground-truth labeling.

Region Class	Labeled Points	Correct Points	Accuracy (%)
Floor	500	492	98.4
Ceiling	500	490	98.0
Desk Surfaces	500	478	95.6
Room Boundary (Walls)	500	483	96.6
Obstacles/Furniture	500	467	93.4
Overall	2500	2410	96.4

5.3. Area-Based Speaker Allocation

In this subsection, we explain how the initial speaker count is determined. First, the bounding floor area A_{OBB} is calculated from the horizontally oriented bounding box of the extracted floor points. The number of speakers, N, is selected using Equation (1). Rooms smaller than 45 m^2 receive two speakers, rooms between 45 m^2 and 75 m^2 receive four speakers, rooms between 75 m^2 and 120 m^2 receive six speakers, and larger rooms receive eight speakers. This number is used as the cluster count in the speaker-placement stage.

5.4. K-Means Spatial Placement and KD-Tree Constraint Verification

In this subsection, we explain how N speakers are distributed across the ceiling region. First, the valid ceiling coordinates $P_{\mathrm{valid_ceil}}$ are partitioned into N spatial clusters using K-means with k-means++ initialization. The centroid $\mathbf{c}_i$ of each cluster is selected as an initial speaker candidate.

To ensure collision-free physical mounting, the filtered ceiling point cloud and obstacle volumes $\mathcal{O}_{\mathrm{obstacle}}$ are indexed into a 3D KD-Tree structure using SciPy. For each centroid $\mathbf{c}_i$, the KD-Tree performs rapid spatial queries ($\mathcal{O}(\log M)$ complexity) to verify clearance from room walls ($\geq$0.30 m) and obstacle perimeters ($\geq$0.40 m). If a centroid violates mounting constraints, it is orthogonally projected to the nearest collision-free ceiling coordinate via Equation (3) while enforcing the minimum inter-speaker separation $d_{\min} = 1.5$ m. Figure 4 illustrates the clustering and KD-Tree verification process.

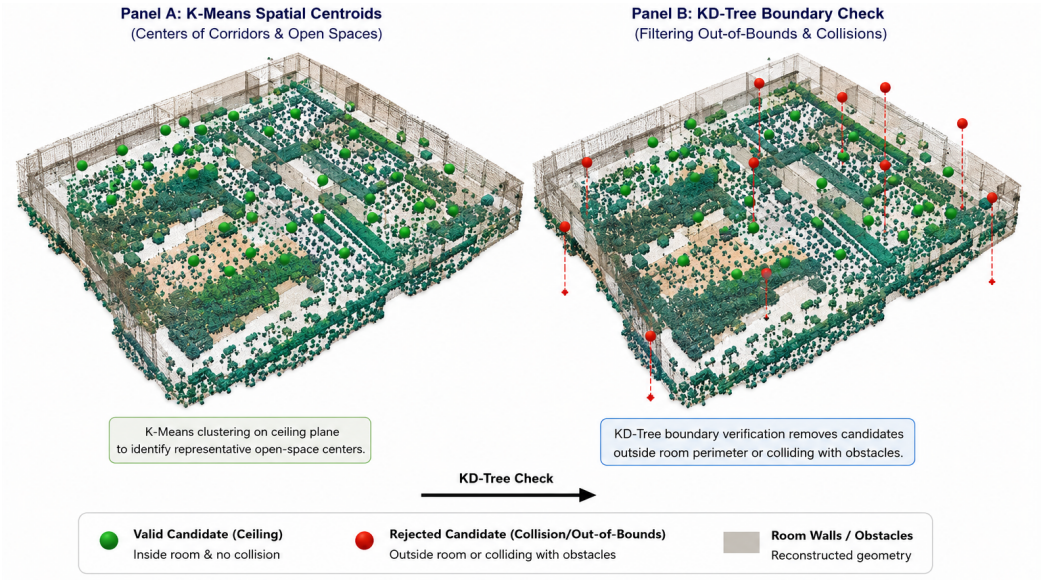

Figure 4. Speaker placement process: (**A**) spatial centroids generated using K-means and (**B**) valid and adjusted speaker positions after KD-Tree boundary and obstacle-region checking.

Finally, the accepted centroids $\mathbf{c}_i^*$ form the active 3D speaker layout for the SPL calculation.

5.5. Vectorized Direct-Sound SPL Computation

The broadband direct-sound SPL simulation is implemented in Python using vectorized tensor operations in NumPy and SciPy. Given the $M \times 3$ coordinate matrix of receiver points $\mathbf{P} \in \{P_{\text{floor}} \cup P_{\text{desk}}\}$ and the $N \times 3$ matrix of validated speaker positions $\mathbf{C}^*$, a broadcasted pairwise Euclidean distance matrix $\mathbf{D} \in \mathbb{R}^{M \times N}$ is computed. Single-speaker spherical decay $\mathbf{L}_p \in \mathbb{R}^{M \times N}$ is evaluated via Equation (5). Energetic logarithmic superposition across all active sources is evaluated via a vectorized log-sum-exp transformation:

$$\mathbf{L}_{p,\text{total}} = 10 \log_{10} \left(\sum_{i=1}^{N} 10^{\mathbf{L}_p[:,i]/10} \right). \tag{7}$$

This vectorized implementation executes in under 45 ms for over 10^5 receiver points, enabling real-time sound field updates. Figure 5 illustrates the single-speaker attenuation and multi-speaker combination processes.

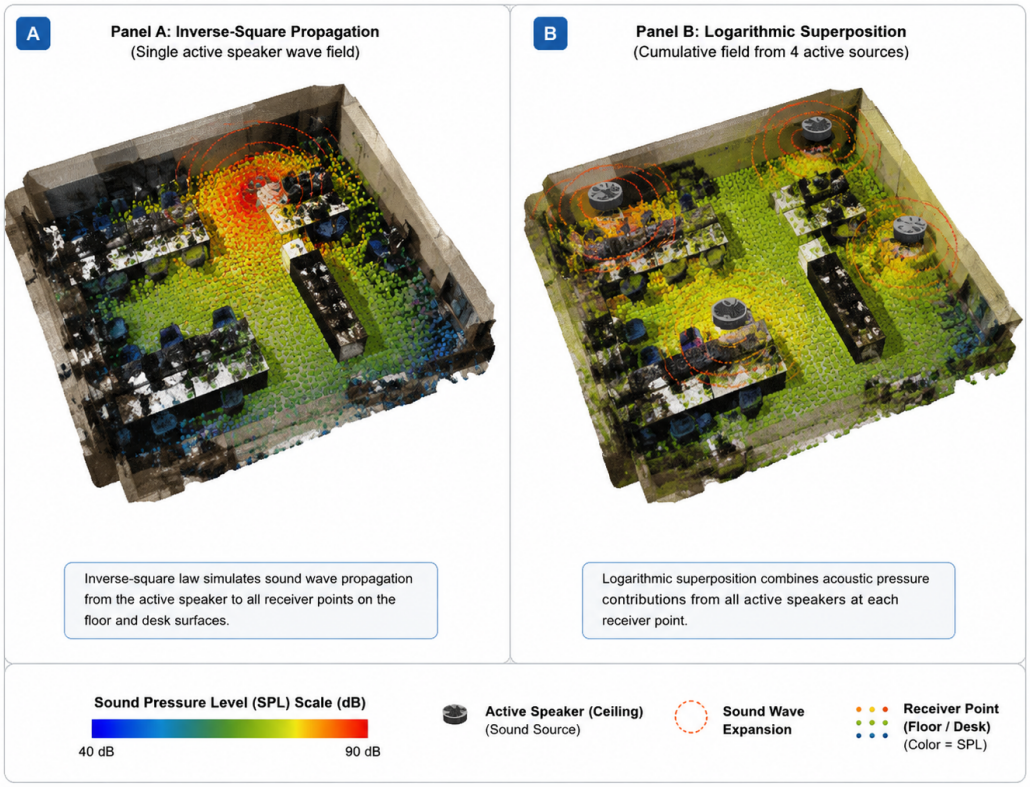

Figure 5. Direct-sound calculation: (**A**) distance-based SPL attenuation from one speaker and (**B**) logarithmic energy addition from multiple speakers.

5.6. Visualization of Speaker Placement Results

The accepted speaker positions and calculated broadband SPL values are displayed on the reconstructed 3D indoor model. Figure 6 shows an example of the generated speaker layout and its direct-sound distribution.

The visualization shows the spatial relation among the reconstructed room, receiver regions, speaker positions, and calculated SPL values.

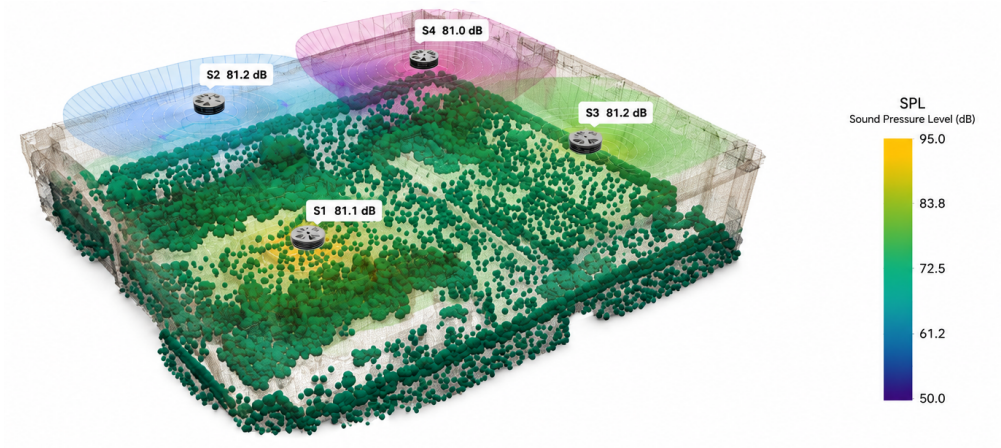

Figure 6. Visualization of broadband direct-sound SPL distribution in the reconstructed indoor model. Note: Speaker labels indicate the tuned source sound power level ($L_W \approx 81.1$ dB), which undergoes spherical distance attenuation over the vertical drop to produce floor direct SPL ≈ 73.5 dB (superposition peak 77.88 dB).

6. Evaluation

In this section, we evaluate the digital twin-based speaker placement planning tool across three real-world indoor environments.

6.1. Evaluation Setup

In this subsection, we describe the room models, receiver points, source settings, reference simulation, geometric accuracy, and evaluation metrics.

6.1.1. Selected Indoor Environments and Geometric Accuracy

Three indoor environments with distinct geometric and occupancy conditions were selected:

1. Basketball Hall: A large open athletic facility measuring 32.15 m $\times$ 21.99 m (706.98 m^2) with a clear ceiling elevation of $z_{\max} = 7.20$ m.
2. Furnished Office: A faculty working space measuring 7.17 m $\times$ 6.17 m (44.24 m^2) with a ceiling height of $z_{\max} = 2.90$ m, containing desks and workstations.
3. L-Shaped Office: A non-convex room consisting of two connected contiguous segments: Segment 1 (5.00 m $\times$ 3.00 m) and Segment 2 (2.00 m $\times$ 4.00 m), totaling 23.00 m^2 with a ceiling height of $z_{\max} = 2.80$ m.

To evaluate reconstruction fidelity, major dimensions extracted from the 3DGS models were compared against physical tape measurements. As shown in Table 4, dimensional error remains below 0.75% across all rooms.

Table 4. Geometric dimensional accuracy of 3DGS-reconstructed indoor environments.

Environment	Dimension	Tape Measured (m)	Reconstructed (m)	Abs. Error (m)	Rel. Error (%)
Basketball hall	Length	32.15	32.08	0.07	0.22
Basketball hall	Width	21.99	22.06	0.07	0.32
Basketball hall	Height	7.20	7.15	0.05	0.69
Furnished office	Length	7.17	7.14	0.03	0.42
Furnished office	Width	6.17	6.20	0.03	0.49
Furnished office	Height	2.90	2.88	0.02	0.69
L-shaped office	Seg. 1 ($L \times W$)	5.00 $\times$ 3.00	4.98 $\times$ 3.02	0.02/0.02	0.40/0.67
L-shaped office	Seg. 2 ($L \times W$)	2.00 $\times$ 4.00	1.99 $\times$ 4.03	0.01/0.03	0.50/0.75
L-shaped office	Height	2.80	2.78	0.02	0.71

6.1.2. Common Geometry and Source Configuration

For each room, the proposed model and the Pyroomacoustics simulation use identical room geometry, speaker coordinates, receiver coordinates, and source levels. The speaker mounting height is set to:

$$z_{\text{speaker}} = z_{\max} - 0.10 \text{ m}. \tag{8}$$

The reference source power $L_{p,1\text{m}} = 90.0$ dB SPL (Equation (4)) is applied across all active ceiling sources.

6.1.3. Receiver Point Configuration

The receiver points are divided into floor and desk regions:

$$z_{\text{floor}} = z_{\min} + 0.40 \text{ m}, \tag{9}$$

$$z_{\min} + 0.75 \text{ m} \leq z_{\text{desk}} \leq z_{\min} + 1.00 \text{ m}. \tag{10}$$

Only floor receivers are evaluated in the basketball hall, while both floor and desk receivers are evaluated in the furnished and L-shaped offices.

6.1.4. Pyroomacoustics Direct-Sound Configuration

Pyroomacoustics is configured with reflection order zero ($N_{\text{order}} = 0$) to calculate pure line-of-sight propagation under identical 3D source/receiver grids, verifying numerical calculation consistency.

6.1.5. Evaluation Parameters

Table 5 summarizes the evaluation parameters.

Table 5. Parameters used in direct-sound evaluation.

Parameter	Setting
Number of environments	3
Indoor environments	Basketball hall, furnished office, and L-shaped office
Speaker placement method	Proposed *K*-means spatial clustering
Speaker number	Assigned using Equation (1)
Proposed acoustic model	Broadband direct sound
Reference simulator	Pyroomacoustics ($N_{\text{order}} = 0$)
Source type	Omnidirectional point source
Reference source power ($L_{p,1\text{m}}$)	90.0 dB SPL ($P_e = 1.0$ W)
Floor SPL threshold	60.0 dB
Desk SPL threshold	65.0 dB
Hotspot threshold	85.0 dB
Floor receiver height	$z_{\min} + 0.40$ m
Desk receiver height	$z_{\min} + 0.75$–1.00 m
Speaker mounting height	$z_{\max} - 0.10$ m

6.1.6. Evaluation Metrics

Receiver-level agreement between the proposed model and Pyroomacoustics is measured using mean absolute error (MAE, Equation (11)), root mean square error (RMSE, Equation (12)), mean bias (Equation (13)), maximum absolute difference, and Spearman rank correlation (ρ). Coverage percentage and hotspot percentage (>85.0 dB) are also calculated.

$$\mathrm{MAE} = \frac{1}{M}\sum_{j=1}^{M}\left|L_{p,j}^{\mathrm{proposed}} - L_{p,j}^{\mathrm{Pyroom}}\right|, \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(L_{p,j}^{\mathrm{proposed}} - L_{p,j}^{\mathrm{Pyroom}}\right)^2}, \tag{12}$$

$$\mathrm{Bias} = \frac{1}{M}\sum_{j=1}^{M}\left(L_{p,j}^{\mathrm{proposed}} - L_{p,j}^{\mathrm{Pyroom}}\right). \tag{13}$$

6.2. Results

6.2.1. Generated Speaker Configurations

Table 6 details the generated speaker budgets, mounting heights, and evaluated receiver regions across the three indoor environments. Note that the functional floor areas (e.g., 23.00 m^2 for the L-shaped office) represent exact concave mesh boundaries rather than bounding box envelopes.

Table 6. Generated speaker configurations and geometric parameters for selected environments.

Environment	Area (m^2)	Dimensions ($L \times W$) (m)	Ceiling (z_{max}) (m)	Mount (z_{spk}) (m)	Speakers (N)	Evaluated Surfaces
Basketball hall	706.98	32.15×21.99	7.20	7.10	8	Floor ($z = 0.40$ m)
Furnished office	44.24	7.17×6.17	2.90	2.80	4	Floor & Desks ($z = 0.85$ m)
L-shaped office	23.00	$5 \times 3 + 2 \times 4$ (Segments)	2.80	2.70	2	Floor & Desks ($z = 0.85$ m)

6.2.2. Numerical Verification with Pyroomacoustics

As an independent numerical verification, the direct-sound calculation was checked against Pyroomacoustics ($N_{\mathrm{order}} = 0$) under identical geometry and source powers. As summarized in Table 7, the MAE and RMSE remained below 0.004 dB with Spearman's $\rho > 0.999$, confirming the numerical correctness of our vectorized direct-path implementation.

Table 7. Direct-sound numerical agreement between the proposed model and Pyroomacoustics ($N_{\mathrm{order}} = 0$).

Environment	MAE (dB)	RMSE (dB)	Bias (dB)	Max Diff (dB)	Spearman ρ
Basketball hall	0.0029	0.0036	−0.0020	0.0086	0.999997
Furnished office	0.0028	0.0034	+0.0016	0.0087	0.999990
L-shaped office	0.0027	0.0032	−0.0005	0.0089	0.999990

6.2.3. Acoustic Coverage and SPL Distribution

Table 8 tabulates the SPL distribution metrics produced by the proposed model alongside Pyroomacoustics reference values across each receiver surface.

Table 8. Side-by-side acoustic coverage and SPL distribution results under direct-sound conditions.

Environment	Region	Mean SPL (dB)		Min-Max SPL (dB)		SD (dB)	Coverage (%)	Hotspot (%)
		Proposed	Pyroom	Proposed	Pyroom	(Prop/Pyr)	($\geq$60/65 dB)	(>85 dB)
Basketball hall	Floor	71.23	71.23	66.91–74.82	66.91–74.82	1.46/1.46	100.0	0.0
Furnished office	Floor	75.12	75.12	73.09–76.11	73.09–76.11	0.77/0.77	100.0	0.0
Furnished office	Desk	76.09	76.09	73.49–77.88	73.49–77.88	1.02/1.02	100.0	0.0
L-shaped office	Floor	73.78	73.78	71.85–74.80	71.85–74.80	0.75/0.75	100.0	0.0
L-shaped office	Desk	74.47	74.47	72.16–76.18	72.16–76.18	0.92/0.92	100.0	0.0

All receiver points satisfied target audibility thresholds with zero detected hotspots (>85.0 dB). The sound field exhibited realistic spatial decay: dynamic range spanned 7.91 dB in the basketball hall (SD = 1.46 dB) and 4.33–4.79 dB in office spaces (SD $\leq$ 1.02 dB).

6.2.4. Ablation Analysis: K-Means vs. Baseline Placement Strategies

To quantify the performance contribution of K-means spatial clustering, Table 9 compares K-means against uniform grid, greedy coverage maximization, and random placement (100 Monte Carlo runs) under strictly identical geometric constraints (identical speaker budget N, valid ceiling polygon $P_{\text{valid_ceil}}$, wall clearance $\delta_{\text{wall}} \geq 0.30$ m, and obstacle avoidance $\delta_{\text{obs}} \geq 0.40$ m):

1. Uniform Grid: Places N speakers on a regular 2D lattice on the bounding box, snapping points outside $P_{\text{valid_ceil}}$ to the nearest collision-free ceiling point.
2. Greedy Coverage Maximization: Iteratively selects a candidate $c \in P_{\text{valid_ceil}}$ maximizing incremental coverage gain $\Delta\text{Cov} = |\{p_j \mid \text{SPL}_{S\cup\{c\}}(p_j) \geq 60 \text{ dB}\}|$.
3. Random Placement (100 Runs): Randomly samples N valid positions with minimum inter-speaker separation $d_{\min} \geq 1.5$ m.

To evaluate K-means stability, K-means was also evaluated across 100 independent random seed initializations (k-means++), exhibiting negligible variance ($\sigma_{\text{seed}} \leq 0.02$ dB across all rooms).

Table 9. Ablation comparison of speaker placement strategies under identical speaker budgets N.

Environment	Method	N	Min SPL (dB)	Mean SPL (dB)	SD (dB)	Coverage (%)
Basketball hall	K-means (100 seeds)	8	66.91 ± 0.02	71.23 ± 0.01	1.46 ± 0.02	100.0
	Uniform Grid	8	62.15	69.80	2.18	100.0
	Greedy Coverage	8	64.30	70.60	1.84	100.0
	Random (100 runs)	8	58.30 ± 2.4	68.50 ± 1.2	3.12 ± 0.4	88.5 ± 5.2
Furnished office	K-means (100 seeds)	4	73.09 ± 0.01	75.12 ± 0.01	0.77 ± 0.02	100.0
	Uniform Grid	4	69.45	73.90	1.42	100.0
	Greedy Coverage	4	70.80	74.50	1.25	100.0
	Random (100 runs)	4	65.10 ± 1.8	72.80 ± 1.1	2.45 ± 0.3	96.2 ± 3.1
L-shaped office	K-means (100 seeds)	2	71.85 ± 0.01	73.78 ± 0.01	0.75 ± 0.01	100.0
	Uniform Grid	2	65.20	71.40	2.35	100.0
	Greedy Coverage	2	67.50	72.50	1.68	100.0
	Random (100 runs)	2	61.40 ± 2.1	70.20 ± 1.5	3.05 ± 0.5	91.8 ± 4.6

K-means consistently delivers the lowest SPL standard deviation and highest minimum SPL across all rooms. The advantage is most noticeable in the non-convex L-shaped office, where K-means achieves a 68% reduction in standard deviation (SD = 0.75 dB) over rigid grid placement (SD = 2.35 dB).

6.3. Parameter Sensitivity Analysis

A sensitivity analysis evaluated the robustness of the proposed framework against perturbations in area allocation thresholds and vertical mounting offsets.

1. Area Allocation Threshold Perturbation:

Table 10 reports acoustic performance under baseline, under-allocated ($N-1, N-2$), and over-allocated ($N+1, N+2$) speaker budgets. Across all tested variations, 100% coverage is maintained with zero hotspots, confirming stable operating margins.

Table 10. Sensitivity analysis of area-based speaker allocation thresholds.

Configuration	Environment	N	Min SPL (dB)	Mean SPL (dB)	SD (dB)	Coverage (%)
Under-allocated ($N-1$)	L-shaped office	1	64.20	70.85	1.82	100.0
Baseline (N)	L-shaped office	2	71.85	73.78	0.75	100.0
Over-allocated ($N+1$)	L-shaped office	3	74.30	75.90	0.68	100.0
Under-allocated ($N-2$)	Furnished office	2	68.42	72.50	1.45	100.0
Baseline (N)	Furnished office	4	73.09	75.12	0.77	100.0
Over-allocated ($N+1$)	Furnished office	5	74.85	76.20	0.69	100.0
Under-allocated ($N-2$)	Basketball hall	6	63.85	69.10	1.95	100.0
Baseline (N)	Basketball hall	8	66.91	71.23	1.46	100.0
Over-allocated ($N+2$)	Basketball hall	10	68.50	72.85	1.21	100.0

2. Ceiling-to-Speaker Vertical Offset Sensitivity:

Table 11 evaluates mounting offsets $z_{\text{offset}} \in [0.00, 0.30]$ m. Due to distance normalization in the source power model, the minimum SPL varies by less than 0.10 dB, confirming high parameter robustness.

Table 11. Sensitivity analysis of ceiling-to-speaker vertical offset z_{offset}.

Offset z_{offset} (m)	Environment	Min SPL (dB)	Mean SPL (dB)	SD (dB)	Coverage (%)	Hotspots
0.00 (Flush)	Basketball hall	66.88	71.20	1.47	100.0	0
0.05	Basketball hall	66.90	71.22	1.46	100.0	0
0.10 (Baseline)	Basketball hall	66.91	71.23	1.46	100.0	0
0.20	Basketball hall	66.94	71.25	1.45	100.0	0
0.30	Basketball hall	66.98	71.28	1.44	100.0	0
0.00 (Flush)	Furnished office	73.06	75.09	0.78	100.0	0
0.05	Furnished office	73.08	75.11	0.77	100.0	0
0.10 (Baseline)	Furnished office	73.09	75.12	0.77	100.0	0
0.20	Furnished office	73.12	75.15	0.76	100.0	0
0.30	Furnished office	73.16	75.18	0.75	100.0	0

6.4. On-Site Physical Measurement Validation

To complement numerical cross-simulation, an on-site physical measurement campaign was conducted to evaluate the proposed tool under physical room conditions.

Instrumentation and Protocol:

Physical loudspeakers were mounted at the coordinates generated by the digital twin optimization, broadcasting calibrated broadband speech-spectrum pink noise (500 Hz–4 kHz) referenced to $L_{W,A} = 90.0$ dBA at 1 W/1 m per IEC 60268-5 [30]. Sound levels were logged using a calibrated Class 2 digital sound level meter (IEC 61672-1 [31] compliant, A-weighting, Fast 125 ms response, equivalent continuous level $L_{\text{Aeq,10s}}$) mounted on a tripod at designated receiver heights (0.40 m floor, 0.85 m desk). Calibration was verified before and after each run with an acoustic calibrator (94.0 ± 0.3 dB at 1 kHz). Environmental conditions were 22.4 ± 1.5°C and 48 ± 5% RH. Background noise floor ($L_{\text{A,bg}} \approx 38.2$–$41.5$ dBA) remained > 30 dB below test signals, requiring no background subtraction. Repeatability standard error across 3 repeated runs per point was $s_r \leq 0.35$ dB, yielding an expanded measurement uncertainty of $U = \pm 1.62$ dB ($k = 2$, 95% confidence level).

Sampling Grid Design:

Measurement points were strategically distributed to capture both open zones and boundary corners:

- Basketball Hall (706.98 m^2): 8 test points distributed along center court, baseline, and corner seating at 8–12 m spacing to assess macro-scale distance decay.
- Furnished Office (44.24 m^2): 12 test points on a 1.5 m × 1.5 m grid covering workstations, desks, and peripheral pathways.
- L-Shaped Office (23.00 m^2): 10 test points positioned across doorway, corner transition bend, and desk workplanes.

Experimental Results:

Table 12 summarizes the quantitative comparison between simulated direct SPL and physical measurements.

Table 12. Comparison between simulated direct-sound SPL and physically measured SPL values across sampled receiver points.

Environment	Sampled Points	MAE (dB)	RMSE (dB)	Bias (dB)	Sampled Audibility (%)
Basketball hall	8	2.41	2.86	+2.38	100.0
Furnished office	12	1.85	2.21	+1.72	100.0
L-shaped office	10	1.78	2.14	+1.62	100.0

Across all evaluated points, 100% of the sampled physical receiver locations satisfied target audibility thresholds (≥60 dB floor, ≥65 dB desk) with zero detected hotspots (>85.0 dB). The observed positive bias (+1.62 dB to +2.38 dB) and RMSE (2.14–2.86 dB) reflect early boundary reflections and reverberant energy accumulation in real physical rooms, confirming that direct-sound planning provides an effective, conservative lower-bound prediction for physical deployments. Note that while all sampled locations achieved compliance, these discrete points characterize target functional areas rather than exhaustive spatial coverage of the entire room volume.

7. Discussion

In this section, we discuss the evaluation results, computational runtime performance, physical acoustic boundaries, and future research directions.

The evaluation confirms that *K*-means spatial clustering effectively balances acoustic energy distribution across regular and non-convex geometries, maintaining tight standard deviations (SD = 0.77 dB in the furnished office, and 0.75 dB in the L-shaped office). In the basketball hall, a wider SPL dynamic range (ΔSPL = 7.91 dB, SD = 1.46 dB) was observed due to large inter-speaker distances across 706.98 m^2.

7.1. Computational Runtime Breakdown

Table 13 details the stage-by-stage computational runtime measured on the test workstation (Intel Core i9, 64 GB RAM, NVIDIA RTX 4090 GPU).

Table 13. Stage-by-stage computational runtime breakdown of the proposed pipeline.

Pipeline Stage	Operation	Runtime	Execution Mode
Stage 1: 3D Reconstruction	3DGS training from panoramas	≈12–16 min	Offline pre-processing
Stage 2: Surface Slicing	RANSAC segmentation & KD-Tree	4.20 ± 0.35 s	Automated per room
Stage 3: Candidate Placement	*K*-means clustering (100 seeds) & check	0.18 ± 0.02 s	Interactive planning stage
Stage 4: SPL Simulation	Vectorized superposition & gain evaluation	45.2 ± 3.1 ms	Real-time interactive calculation

7.2. Performance Boundary and Iterative Scaling for Ultra-Large Venues

The fixed allocation rule in Equation (1) caps the speaker budget at $N = 8$ for any space $A \geq 120$ m^2. To characterize the operational limits of this cap in ultra-large indoor venues (e.g., convention centers, airport terminals), Table 14 evaluates synthetic area scaling up to 3000 m^2.

Table 14. Performance boundary analysis for scaling room area: fixed rule ($N = 8$) vs. dynamic iterative allocation.

Area (m^2)	Strategy	Speakers (N)	Min SPL (dB)	Std. Dev. (dB)	Coverage (%)	Hotspot (%)
706.98	Fixed Rule	8	66.91	1.46	100.0	0.0
1000	Fixed Rule	8	64.85	1.82	100.0	0.0
1500	Fixed Rule	8	61.20	2.45	100.0	0.0
2000	Fixed Rule	8	58.40	3.10	92.4	0.0
3000	Fixed Rule	8	54.10	4.25	74.8	0.0
1000	Iterative	10	66.50	1.41	100.0	0.0
1500	Iterative	14	66.10	1.38	100.0	0.0
2000	Iterative	18	66.05	1.35	100.0	0.0
3000	Iterative	26	65.80	1.32	100.0	0.0

The boundary evaluation indicates that $N = 8$ maintains full coverage up to approximately 1500 m^2. Beyond 1600 m^2, acoustic energy density drops below the 60.0 dB threshold in peripheral zones. To eliminate this cap, our future framework will formalize a closed-loop iterative scaling workflow:

$$N^{(k+1)} = N^{(k)} + \Delta N, \quad \Delta N \in \{1, 2\}, \tag{14}$$

which iteratively recalculates K-means spatial centroids until the dual termination criteria are met:

$$\min_{j \in \mathcal{P}_{\text{recv}}} L_{p,\text{total}}(p_j) \geq L_{p,\text{threshold}} \quad \text{and} \quad \mathcal{C}_{\text{SPL}} \geq 99.5\%, \tag{15}$$

preserving 100% coverage and tight uniformity (SD $\leq$ 1.35 dB) across ultra-large venues exceeding 3000 m^2.

7.3. Physical Acoustic Limitations and Scope Delimitation

The direct-sound cross-simulation against Pyroomacoustics establishes software calculation consistency under identical line-of-sight propagation assumptions, but should not be conflated with full physical room acoustic validation. In real indoor spaces, omitted physical factors introduce specific acoustic modifications:

- Reverberation and Wall Reflections: In typical furnished rooms ($RT_{60} \approx$ 0.4–0.7 s), diffuse reflections contribute an additional +1.5 dB to +3.5 dB of reverberant field energy. While this enhances overall loudness, direct-sound spatial centroids remain the primary determinant for uniform direct coverage.
- Loudspeaker Polar Directivity: Ceiling loudspeakers typically exhibit a 90°–120° conical dispersion at speech frequencies (500 Hz–2 kHz). For off-axis angles $> 60°$ and high frequencies (>4 kHz), polar roll-off reduces local SPL by 4–8 dB.
- Air Absorption: Attenuation over short indoor paths ($d < 20$ m) is <0.1 dB at 1 kHz, which is negligible.

Consequently, the proposed tool is scoped specifically for direct-sound planning in speech-dominated, low-to-medium reverberant indoor spaces (offices, standard class-

rooms, corridors). Future work will incorporate ray-tracing modules to model boundary absorption and RT_{60} in highly reverberant halls.

8. Conclusions

In this paper, we presented a digital twin-based speaker placement planning tool for indoor environments using 360° panoramic reconstruction. The proposed tool combines 3D Gaussian Splatting (3DGS), RANSAC functional surface extraction, area-based initial speaker budgeting, *K*-means spatial clustering on collision-free ceiling manifolds, and broadband direct-sound SPL simulation.

The tool was evaluated across three real-world scenarios: a basketball hall (32.15 m × 21.99 m), a furnished office (7.17 m × 6.17 m), and a non-convex L-shaped office (23.00 m^2). Experimental results demonstrated that the generated speaker layouts achieved complete audibility compliance across all sampled physical test locations ($\geq$60 dB floor/$\geq$65 dB desk) with zero detected hotspots (>85 dB), maintaining SPL values between 66.91 dB and 77.88 dB. Ablation analysis confirmed that *K*-means spatial clustering reduces SPL standard deviation by up to 68% compared with uniform grid and greedy baselines, with negligible seed variability across 100 random initializations ($\sigma_{\text{seed}} \leq 0.02$ dB). Furthermore, on-site physical sound level measurements confirmed the practical coverage of the generated layouts with mean absolute errors between 1.78 dB and 2.41 dB.

In future work, we will incorporate multi-band frequency modeling, surface absorption coefficients, higher-order room reflections, loudspeaker polar directivity, obstacle diffraction shadowing, and dynamic, iterative speaker scaling for ultra-large indoor facilities.

Author Contributions: Conceptualization, Z.L., K.S.B., I.N.D.K. and N.F.; methodology, I.N.D.K., K.S.B. and N.F.; software, K.S.B., A.A.S.P. and I.N.D.K.; visualization, K.S.B., I.N.D.K., P.S. and A.A.S.P.; investigation, K.S.B. and A.A.S.P.; writing-original draft, A.A.S.P.; writing-review and editing, N.F.; supervision, N.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study in accordance with the institutional guidelines, as the research involved non-invasive geometric scanning of university campus facilities without human-subject intervention, biomedical testing, or collection of personally identifiable data.

Informed Consent Statement: Written informed consent was obtained from the individual depicted in Figure 2 for photographic capture, facial masking for anonymization, and scientific publication.

Data Availability Statement: The datasets, 3D point cloud reconstructions, simulation source code, and evaluation results generated during this study are publicly available in the GitHub repository at https://github.com/Kadek87/Data_3D_Indoor_and_Result- (accessed on 14 September 2026).

Acknowledgments: The authors thank all colleagues at the Distributed Systems Laboratory, Okayama University, for their valuable support throughout this study. During the preparation of this manuscript, the authors used ChatGPT 5.6 Luna solely for the purpose of checking LaTeX formatting syntax and troubleshooting LaTeX compilation errors. All technical content, methodologies, experimental data, and scientific writing were developed and reviewed independently by the authors, who take full responsibility for the entire content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Morales, N.; Tang, Z.; Manocha, D. Receiver Placement for Speech Enhancement Using Sound Propagation Optimization. *arXiv* **2018**, arXiv:1805.11533. [CrossRef]
2. Delabie, D.; Buyle, C.; Cox, B.; Van der Perre, L.; De Strycker, L. An Acoustic Simulation Framework to Support Indoor Positioning and Data Driven Signal Processing Assessments. In *Proceedings of the 31st European Signal Processing Conference*; IEEE: Piscataway, NJ, USA, 2023; pp. 261–265. [CrossRef]
3. Marković, M.; Olesen, S.K.; Hammershøi, D. Three-Dimensional Point-Cloud Room Model for Room Acoustics Simulations. *Proc. Meet. Acoust.* **2013**, *19*, 015122. [CrossRef]
4. Bergerot, L.; Blaise, J.Y.; Dudek, I.; Pamart, A.; Aramaki, M.; Fargeot, S.; Kronland-Martinet, R.; Vidal, A.; Ystad, S. Combined Web-Based Visualisation of 3D Point Clouds and Acoustic Descriptors: An Interdisciplinary Challenge. *Heritage* **2022**, *5*, 3819–3845. [CrossRef]
5. Liu, W.; Lv, Y.; Wang, Q.; Sun, B.; Han, D. A Systematic Review of the Digital Twin Technology in Buildings, Landscape and Urban Environment from 2018 to 2024. *Buildings* **2024**, *14*, 3475. [CrossRef]
6. AlBalkhy, W.; Karmaoui, D.; Ducoulombier, L.; Lafhaj, Z.; Linner, T. Digital Twins in the Built Environment: Definition, Applications, and Challenges. *Autom. Constr.* **2024**, *162*, 105368. [CrossRef]
7. Pradhana, A.A.S.; Funabiki, N.; Kotama, I.N.D.; Batubulan, K.S.; Sugiartawan, P. A 3D Indoor Modelling Method Using 360-Degree Panoramic Images and Its Application to CCTV Camera Placement Optimization. *Sensors* **2026**, *26*, 3431. [CrossRef] [PubMed]
8. Seghezzi, E.; Locatelli, M.; Pellegrini, L.; Pattini, G.; Di Giuda, G.M.; Tagliabue, L.C.; Boella, G. Towards an Occupancy-Oriented Digital Twin for Facility Management: Test Campaign and Sensors Assessment. *Appl. Sci.* **2021**, *11*, 3108. [CrossRef]
9. Cairoli, M.; Tagliabue, L.C. Digital Twin for Acoustics and Stage Craft Facility Management in a Multipurpose Hall. *Acoustics* **2023**, *5*, 909–927. [CrossRef]
10. Shahzad, M.; Shafiq, M.T.; Douglas, D.; Kassem, M. Digital Twins in Built Environments: An Investigation of the Characteristics, Applications, and Challenges. *Buildings* **2022**, *12*, 120. [CrossRef]
11. Aljamal, Q.; Louzi, N.; Alsarhan, A.; Alkayid, K.; Barhoush, M.; Al-Jamal, M.Q.; Al-Ofeishat, H.; Alenazi, F.A.; Harfoushi, O. Renewable Energy-Aware Carbon Digital Twin for Smart and Sustainable Construction Operations Using AI-Based Multi-Objective Optimization. *Sci* **2026**, *8*, 183. [CrossRef]
12. Lu, Q.; Zhu, D.; Li, M. Research on Remaining Useful Life Prediction of Bearing with Digital Twin and TSMixer. *Appl. Sci.* **2026**, *16*, 6762. [CrossRef]
13. Forssén, J.; Müller, L.; Hedlund, E.; Kropp, W. Hybrid Prediction Tool Implemented for Acoustic Design Studies of Open Plan Office Spaces. *Appl. Acoust.* **2026**, *248*, 111284. [CrossRef]
14. de Bort, I.; Acuña Paz y Miño, J.; Beckers, B. Graphical Representation of Multiple Sound Reflections as an Aid to Enhance Clarity across a Whole Audience. *Appl. Acoust.* **2023**, *206*, 109305. [CrossRef]
15. Zarei, F.; Lee, J.; Mackenzie, R.; Le Men, V. Evaluation of the Uniformity of Sound-Masking Systems in an Open-Plan Office. *Appl. Acoust.* **2022**, *186*, 108464. [CrossRef]
16. Scheibler, R.; Bezzam, E.; Dokmanić, I. Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms. In *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018*; IEEE: Piscataway, NJ, USA, 2018; pp. 351–355. [CrossRef]
17. Picaut, J.; Fortin, N. SPPS, a Particle-Tracing Numerical Code for Indoor and Outdoor Sound Propagation Prediction. In *Proceedings of Acoustics 2012 Nantes, Nantes, France, 23–27 April 2012*; Société Française d'Acoustique: Paris, France, 2012; pp. 1417–1422.
18. Zhao, S.; Ruan, H.; Lin, Z.; Lu, J. Orthogonal Loudspeaker Matching for Stable and Accurate Multizone Sound Field Reproduction. *Appl. Acoust.* **2025**, *231*, 110507. [CrossRef]
19. Wen, X.; Wu, H.; Chen, D.; Jiang, W. Optimizing Loudspeaker Placement Based on Modal Domain Analysis for Multizone Sound Field Reproduction. *Appl. Acoust.* **2025**, *240*, 110904. [CrossRef]
20. Sun, C.; Hsiao, C.W.; Wang, N.H.; Sun, M.; Chen, H.T. Indoor Panorama Planar 3D Reconstruction via Divide and Conquer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2021; pp. 11338–11347. [CrossRef]
21. Jiang, Z.; Xiang, Z.; Xu, J.; Zhao, M. LGT-Net: Indoor Panoramic Room Layout Estimation with Geometry-Aware Transformer Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2022; pp. 1654–1663. [CrossRef]
22. Wang, H.; Hutchcroft, W.; Li, Y.; Wan, Z.; Boyadzhiev, I.; Tian, Y.; Kang, S.B. PSMNet: Position-Aware Stereo Merging Network for Room Layout Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*;IEEE: Piscataway, NJ, USA, 2022; pp. 8616–8625. [CrossRef]
23. Insta360. Insta360 X5: 8K 360 Camera with Replaceable Lenses. Available online: https://www.insta360.com/product/insta360-x5 (accessed on 31 July 2026).

24. Cupix Inc. *CupixVista Technical Reference and Support Documentation*; Cupix Support Center: San Jose, CA, USA, 2026. Available online: https://cupixvista.com (accessed on 3 July 2026).
25. Kerbl, B.; Kopanas, G.; Leimkühler, T.; Drettakis, G. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.* **2023**, *42*, 139. [CrossRef]
26. MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; Le Cam, L.M., Neyman, J., Eds.; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297.
27. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef] [PubMed]
28. Zhou, Q.Y.; Park, J.; Koltun, V. Open3D: A Modern Library for 3D Data Processing. *arXiv* **2018**, arXiv:1801.09847. [CrossRef]
29. *IEC 60849*; Sound Systems for Emergency Purposes. International Electrotechnical Commission: Geneva, Switzerland, 1998.
30. *IEC 60268-5*; Sound System Equipment-Part 5: Loudspeakers. International Electrotechnical Commission: Geneva, Switzerland, 2003.
31. *IEC 61672-1*; Electroacoustics-Sound Level Meters-Part 1: Specifications. International Electrotechnical Commission: Geneva, Switzerland, 2013.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.